THE ROYAL SOCIETY

doi 10.1098/rspb.2000.1397

High guanine–cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes

Laurence D. Hurst[1]* **and Alexa R. Merchant**[2]

[1]*Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK* (*l.d.hurst@bath.ac.uk*)
[2]*Stonar School, Cottles Park, Atworth, Melksham, Wiltshire SN12 8NT, UK* (*alexa_merchant@yahoo.com*)

The causes of the variation between genomes in their guanine (G) and cytosine (C) content is one of the central issues in evolutionary genomics. The thermal adaptation hypothesis conjectures that, as G:C pairs in DNA are more thermally stable than adenonine:thymine pairs, high GC content may be a selective response to high temperature. A compilation of data on genomic GC content and optimal growth temperature for numerous prokaryotes failed to demonstrate the predicted correlation. By contrast, the GC content of structural RNAs is higher at high temperatures. The issue that we address here is whether more freely evolving sites in exons (i.e. codonic third positions) evolve in the same manner as genomic DNA as a whole, showing no correlated response, or like structural RNAs showing a strong correlation. The latter pattern would provide strong support for the thermal adaptation hypothesis, as the variation in GC content between orthologous genes is typically most profoundly seen at codon third sites (GC_3). Simple analysis of completely sequenced prokaryotic genomes shows that GC_3, but not genomic GC, is higher on average in thermophilic species. This demonstrates, if nothing else, that the results from the two measures cannot be presumed to be the same. A proper analysis, however, requires phylogenetic control. Here, therefore, we report the results of a comparative analysis of GC composition and optimal growth temperature for over 100 prokaryotes. Comparative analysis fails to show, in either Archea or Eubacteria, any hint of connection between optimal growth temperature and GC content in the genome as a whole, in protein-coding regions or, more crucially, at GC_3. Conversely, comparable analysis confirms that GC content of structural RNA is strongly correlated with optimal temperature. Against the expectations of the thermal adaptation hypothesis, within prokaryotes GC content in protein-coding genes, even at relatively freely evolving sites, cannot be considered an adaptation to the thermal environment.

Keywords: GC content; isochores; genome evolution; thermal adaptation

1. INTRODUCTION

There exists considerable variation between species in the guanine (G) and cytosine (C) content of their genes. The causes of this variation is one of the central issues in evolutionary genomics and an important focus for the debate between neutralists and selectionists. Some selectionist models (Bernardi 2000; Bernardi & Bernardi 1986; Salinas *et al.* 1988) posit a link between GC content and temperature. G:C pairs are more thermally stable than adenine (A) and thymine (T) pairs (Wada & Suyama 1986), G:C pairs being connected by three hydrogen bonds and A:T pairs by two. It has, therefore, been conjectured that the higher GC content of birds and mammals, compared with reptiles and amphibians, may be the result of selection favouring increased GC content, as vertebrates changed from being cold-blooded to warm-blooded (Bernardi 2000; Bernardi & Bernardi 1986). This may be referred to as the homeothermy hypothesis or, more generally, the thermal adaptation hypothesis. A similar suggestion has been made for the evolution of plant genomes (Salinas *et al.* 1988).

An obvious test of such hypotheses is to examine the GC content of prokaryotic genomes, as within this group of organisms there is extensive variation in thermal habit. Some living at deep-sea vents have optimal growth temperatures (T_{opt}) over 100 °C, while others grow best around water's freezing point. A recent analysis (Galtier & Lobry 1997) failed, however, to demonstrate any correlation between genomic GC content and T_{opt} (see also Muto & Osawa 1987), thus appearing to reject the thermal adaptation hypothesis (although no allowance for phylogenetic non-independence was made). However, this report also concluded that the GC content of relatively freely evolving functional RNAs (e.g. 16S and 23S RNA) does covary with temperature. What remains unknown is whether at relatively freely evolving sites in exons (i.e. codon third sites), the GC content (GC_3) is related to temperature.

This is an issue of some substance, as the GC variation between orthologous genes (e.g. those of mammals and amphibians) is most profoundly seen at the third sites and this appears to reflect the GC content of flanking non-coding DNA (Clay *et al.* 1996). The GC content at the first two sites of the codons are very much more tightly constrained and show a less profound correlation with the GC content of flanking DNA. We can then presume that in vertebrates GC_3 is largely affected by the same processes that act on non-coding DNA. In bacteria most DNA is coding, so to test the thermal adaptation hypothesis it is most desirable to consider GC_3 rather than GC of the complete coding DNA or complete genomic DNA.

Limited data point to the possibility that GC_3 may be related to temperature and that it may therefore behave

*Author for correspondence.

Received 7 August 2000 *Accepted* 2 October 2000

Table 1. *GC_{coding} and GC_3 in bacterial genomes that have been completely sequenced (complete list taken from TIGR on 12 October 2000)*

species	GC_{coding} (%)	GC_3 (%)	thermophilic
Aeropyrum pernix[a]	57.50	66.40	yes
Archaeoglobus fulgidus[a]	49.37	58.42	yes
Methanobacterium thermoautotrophicum[a]	50.46	56.59	yes
Methanococcus jannaschii[a]	31.84	24.74	yes
Pyrococcus abyssi[a]	45.16	50.31	yes
Pyrococcus horikoshii[a]	42.32	42.97	yes
Aquifex aeolicus[b]	43.58	47.93	yes
Bacillus subtilis[b]	44.32	44.61	no
Borrelia burgdorferi[b]	29.31	20.82	no
Campylobacter jejuni[b]	32.82	18.96	no
Chlamydia muridarum[b]	39.13	29.92	no
Chlamydia pneumoniae[b]	41.30	34.88	no
Chlamydia trachomatis[b]	41.61	34.30	no
Deinococcus radiodurans[b]	65.72	84.02	no
Escherichia coli[b]	51.37	54.90	no
Haemophilus influenzae[b]	38.76	29.09	no
Helicobacter pylori[b]	39.56	41.95	no
Mycobacterium tuberculosis[b]	65.81	79.67	no
Mycoplasma genitalium[b]	31.64	23.01	no
Mycoplasma pneumoniae[b]	41.05	42.08	no
Neisseria meningitidis[b]	50.14	55.49	no
Rickettsia prowazekii[b]	30.59	18.47	no
Synechocystis sp.[b]	48.66	49.99	no
Thermotoga maritima[b]	46.45	52.62	yes
Treponema pallidum[b]	52.52	54.10	no
Ureaplasma urealyticum[b]	35.20	16.97	no
Vibrio cholerae[b]	47.17	49.08	no

[a] Archeal species.
[b] Eubacterial species.

more like GC contents of structural RNA than of coding sequences as a whole or the whole genome GC. For example, early reports (Kagawa *et al.* 1984; Winter *et al.* 1983) suggested that bacteria that live in hot conditions have a high GC_3. Further, we can divide the completely sequenced prokaryotic genomes into those of thermophilic ($n = 8$) and non-thermophilic species ($n = 19$) (table 1). The two sets are very similar as regards the overall GC content of coding regions. The mean GC content of the eight completely sequenced thermophilic prokaryotes is 45.83 ± 2.62%, which is not significantly different (Mann–Whitney U-test, $p > 0.05$) from 43.51 ± 2.41% for the non-thermophilic species. However, analysis of the more freely evolving GC_3 appears to tell a different story. The mean GC_3 content in the thermophilics is 50.00 ± 4.40% against 41.17 ± 4.44% for the non-thermophilics. The difference is on the cusp of being statistically significant (Mann–Whitney U-test, one-tailed, $p = 0.05$). Two more thermophilic species' genomes are in the process of being completed (*Bacillus stearothermophilus* and *Thermus thermophilus*). Inclusion of the available data from these two (382 and 233 protein-coding sequences, respectively) strengthens the effect (Mann–Whitney U-test, one-tailed, $p = 0.01$). These results additionally suggest that conclusions based on genomic GC, or GC of complete coding sequences, cannot necessarily be assumed to apply to GC_3.

The above simple analysis, like all previous analyses, makes no allowance for phylogenetic non-independence (Harvey & Purvis 1991), which may well be a problem as many thermophiles are archeans. Here then we ask whether, as predicted by the thermal adaptation hypothesis, a correlated response with temperature is a general pattern of more freely evolving sites, i.e. GC_3 and GC of structural RNAs. We also enquire as to whether all previous results are robust to phylogenetic control.

2. METHODS

Archea and Eubacteria have different DNA biologies. Importantly, Archea have histone-like proteins involved in some sort of chromatin structuring (Grayling *et al.* 1996). These may well minimize the tendency for DNA to thermally degrade (Grayling *et al.* 1996). By contrast, Eubacteria do not have histone-like proteins. It may conceivably be the case that a relationship between GC and T_{opt} is found in Eubacteria but not in Archea. We therefore analyse the two separately.

Release 8 of the Ribosomal Database Project II (Maidak *et al.* 2000) provides a small subunit ribosomal RNA (rRNA) phylogeny of 63 species of Archea and 165 species of Eubacteria (www.cme.msu.edu/RDP/html/download.html). These were constructed using weighted neighbour joining (Bruno *et al.* 2000). This method has the advantage of being relatively easily applicable to large data sets while still recovering quantitatively and qualitatively similar trees to maximum-likelihood methods. The method appears to be relatively immune to the problem of long-branch attraction, observed with neighbour joining and parsimony (Bruno *et al.* 2000). While these trees do not represent full knowledge of bacterial rDNA, they are maximally robust in that sequences shorter than 1400 bases and/or with more than 4% of the bases being ambiguous are excluded.

Table 2. *Results of comparative analysis of T_{opt} and GC_{coding}, of GC_3 content in these sequences, of $GC_{genomic}$ and three structural RNAs (16S rDNA, 23S rDNA and tRNA) within the Archea*

(Figures in square brackets are those for an edited data set in which species for which fewer than ten protein-coding genes have been sequenced have been eliminated.)

rooted linear regression of contrasts	number of contrasts	p-value on slope	r^2
$GC_{coding} = -0.08\ T_{opt}$	28	0.50	0.02
$[GC_{coding} = -0.07\ T_{opt}]$	[22]	[0.64]	[0.01]
$GC_3 = -0.16\ T_{opt}$	28	0.55	0.01
$[GC_3 = -0.12\ T_{opt}]$	[22]	[0.71]	[0.01]
$GC_{genomic} = -0.01\ T_{opt}$	24	0.97	0.00
$GC_{RNA16S} = 0.14\ T_{opt}$	45	< 0.0001	0.57
$GC_{RNA23S} = 0.17\ T_{opt}$	17	< 0.0001	0.76
$GC_{tRNA} = 0.16\ T_{opt}$	13	0.0076	0.46

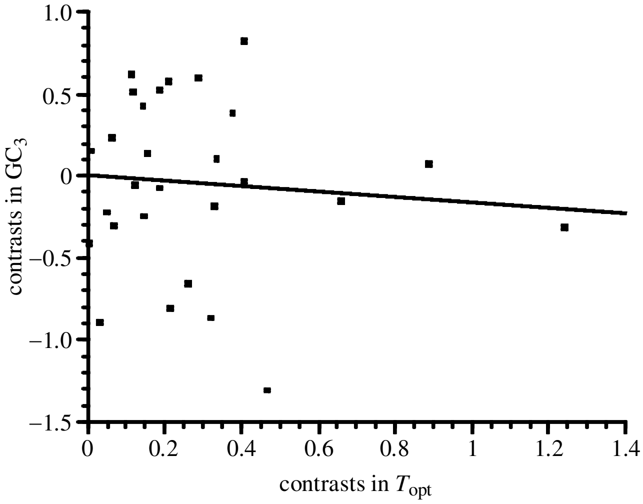

Figure 1. Contrasts in GC_3 as a function of contrasts in T_{opt} from analysis of Archea.

For each of the species identified in the phylogeny we endeavoured to find data on T_{opt}, GC content of all sequenced protein-coding genes (GC_{coding}), GC_3, and genomic GC content ($GC_{genomic}$) assayed by the methods used by Galtier & Lobry (1997). Additionally, GC content for rRNA (16S and 23S) and one transfer RNA (tRNA) for each species was searched for. If a sequence existed (which for 16S it does for all species in the phylogeny) the GC content was estimated by accessing the GenBank files for the gene for each species and splicing out the RNA according to the annotations. When the genome had been completely sequenced and an individual GenBank entry was not available we used the annotations provided at The Institute for Genomic Research (TIGR) (www.tigr.org/).

Data on T_{opt} were obtained mainly from the data set described by Galtier & Lobry (1997) (ftp://biom3.univ-lyon1.fr/pub/datasets/JME97/). Additional data on T_{opt} were obtained primarily from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) (www.dsmz.de/species/strains.htm) but failing this from ATCC (phage.atcc.org/searchengine/ba.html) and from the primary literature. The temperatures reported in DSMZ and American Type Culture Collection (ATCC) agree closely with those used by Galtier & Lobry (1997).

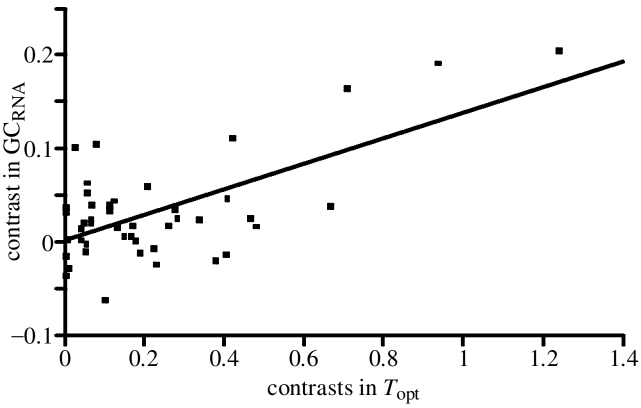

Figure 2. Contrasts in GCRNA as a function of contrasts in T_{opt} from analysis of Archea.

To ensure consistency with the previous analysis, we also employed the data on genomic GC assembled by Galtier & Lobry (1997). GC content of sequenced protein-coding genes (GC_3 and GC_{coding}) was obtained from the Codon Usage Database (www.kazusa.or.jp/codon/). In the final data set we have estimates of both GC content, by at least one of the measures, and T_{opt} for 34 phylogenetically defined archeal species and for 64 species of Eubacteria. Additionally, as within the eubacterial phylogeny each genus was represented only once we could augment our data set by including data from an alternative species within the same genus where no data, or limited data, existed for the one phylogenetically defined. The augmented data set contains information on 74 species. The phylogenies and species employed are available as online supplementary material.

To control for phylogenetic non-independence we use the method of comparative analysis by independent contrasts (CAIC, v. 2.6.2) (Purvis & Rambaut 1995). By taking contrasts between species/nodes we can ask whether the magnitude of the difference in T_{opt} between two species/nodes is reflected in a difference of comparable relative magnitude in GC content. This method allows us to ask directly, therefore, whether on adaptation to a new thermal environment the GC content always shows a correlated response. In all cases we report regressions rooted through the origin. Not rooting does not affect our conclusions.

The phylogenies employed are illustrated in electronic Appendix A available on The Royal Society's web site.

3. RESULTS

Taken across all species, the figures for GC_{coding} and $GC_{genomic}$ are very highly correlated ($r^2 = 0.933$, $n = 56$). Examining the regression of the square of the difference between these two values and the number of sequences obtained per species, shows that only when the number of sequences sampled is very small, is there a larger difference than expected by chance (more than two standard residuals). Only three out of 56 data points were beyond two standard residuals. To be cautious, we additionally analyse all data sets with the data from species with fewer than ten sequences sampled removed, so as to exclude any biases from undersampling.

(a) *Archea*

Phylogenetically uncontrolled analysis reports no correlation between GC and T_{opt}, although if anything the relationship is opposite to that expected: GC_{coding}

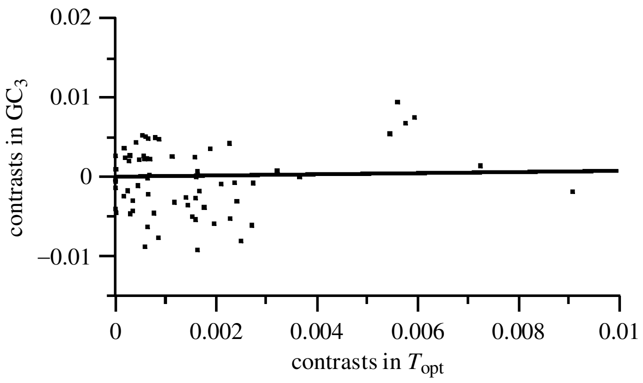

Figure 3. Contrasts in GC$_3$ as a function of contrasts in T_{opt} from analysis of the augmented set of Eubacteria.

Table 3. *Results of comparative analysis of the non-augmented data set of T_{opt} and GC content (by three measures) within the Eubacteria*

(Figures in square brackets are those for an edited data set in which species for which fewer than ten protein-coding genes have been sequenced have been eliminated.)

rooted linear regression of contrasts	number of contrasts	p-value on slope	r^2
$GC_{coding} = 0.02\ T_{opt}$	52	0.90	0.00
$[GC_{coding} = -0.04\ T_{opt}]$	[36]	[0.81]	[0.00]
$GC_3 = 0.11\ T_{opt}$	52	0.75	0.00
$[GC_3 = -0.05\ T_{opt}]$	[36]	[0.88]	[0.00]
$GC_{genomic} = -0.01\ T_{opt}$	39	0.96	0.00
$GC_{RNA16S} = 0.07\ T_{opt}$	120	0.013	0.05
$GC_{RNA23S} = 0.24\ T_{opt}$	17	> 0.0001	0.59
$GC_{tRNA} = 0.05\ T_{opt}$	13	0.53	0.03

$= 60.6 - 0.129\ T_{opt}$, $r^2 = 0.072$, $p = 0.087$, $n = 29$; $GC_3 = 73.4 - 0.209\ \ 129\ T_{opt}$, $r^2 = 0.034$, $p = 0.17$, $n = 29$; $GC_{genomic} = 56.4 - 0.124\ T_{opt}$, $r^2 = 0.009$, $p = 0.28$, $n = 25$. More importantly, we examined contrasts of optimal temperature against all three measures of GC content, but observed no suggestion even that the GC content and growth temperature may be related (table 2). Most importantly, contrasts in GC$_3$ show no covariation with contrasts in T_{opt} (figure 1). Neither this result nor that for GC$_{coding}$ is affected by the removal from the analysis of species in which fewer than ten coding sequences have been studied (table 2). Additionally, we have replaced some species by members from the same genus for which we have better data (a total of four changes, three replacements and one addition). This does not qualitatively affect the results.

By contrast, the GC content of the structural RNAs shows a very pronounced trend with temperature: regression of raw data, $GC_{RNA16S} = 49.4 + 0.183\ T_{opt}$, $r^2 = 0.87$, $p < 0.0001$, $n = 46$; $GC_{RNA23S} = 48 + 0.195\ T_{opt}$, $r^2 = 0.84$, $p < 0.0001$, $n = 18$; $GC_{tRNA} = 55.6 + 0.18\ T_{opt}$, $r^2 = 0.69$, $p < 0.0001$, $n = 14$. The effects remains extremely strong when phylogenetic non-independence is allowed for (table 2). The contrasts for the largest of the analyses (16S) are shown in figure 2.

(b) *Eubacteria*

Phylogenetically uncontrolled analysis within the Eubacteria also reports no correlation between GC and T_{opt}, and again, if anything the relationship is opposite to that expected, with higher temperatures being associated with lower GC: $GC_{coding} = 53.5 - 0.05\ T_{opt}$, $r^2 = 0.00$, $p = 0.689$, $n = 66$; $GC_3 = 58.9 - 0.032129\ T_{opt}$, $r^2 = 0.00$, $p = 0.89$, $n = 66$; $GC_{genomic} = 52.1 + 0.098\ T_{opt}$, $r^2 = 0.00$, $p = 0.57$, $n = 55$.

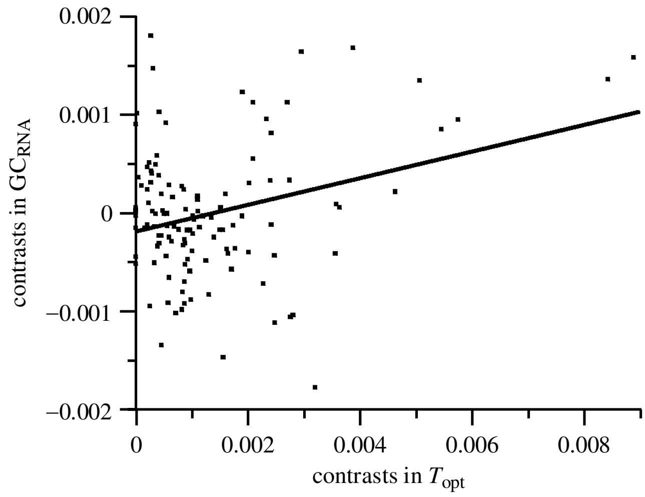

Figure 4. Contrasts in GC$_{RNA}$ as a function of contrasts in T_{opt} from analysis of the augmented set of Eubacteria.

Table 4. *Results of comparative analysis of the augmented data set of T_{opt} and GC content (by three measures) within the Eubacteria*

(Figures in square brackets are those for an edited data set in which species for which fewer than ten protein-coding genes have been sequenced have been eliminated.)

rooted linear regression of contrasts	number of contrasts	p-value on slope	r^2
$GC_{coding} = 0.02\ T_{opt}$	72	0.89	0.00
$[GC_{coding} = 0.08\ T_{opt}]$	[49]	[0.57]	[0.01]
$GC_3 = 0.06\ T_{opt}$	72	0.80	0.00
$[GC_3 = 0.11\ T_{opt}]$	[49]	[0.70]	[0.00]
$GC_{genomic} = 0.01\ T_{opt}$	52	0.94	0.00

For the comparative analysis within the Eubacteria we analysed a non-augmented data set, in which each data point is defined for the species that is phylogenetically defined, and an augmented data set in which a representative of the genus was used, but not necessarily that for which the phylogeny was defined. Within the non-augmented data set, as within the Archea, we found no evidence for correlated evolution of GC content and T_{opt} (table 3). As previously, neither the result for GC$_3$ nor that for GC$_{coding}$ is affected by the removal from the analysis of species in which fewer than ten coding sequences have been studied (table 3). Analysis of the augmented data set provides the same results (for contrasts of GC$_3$ versus T_{opt}, see figure 3). Again, neither the result for GC$_3$ nor that for GC$_{coding}$ is affected by the removal from the analysis of species in which fewer than ten coding sequences have been studied (table 4).

As with the Archea, the GC content of the 16S and 23S RNAs is higher at higher temperatures. Regression of raw data, $GC_{RNA16S} = 50.4 + 0.145\ T_{opt}$, $r^2 = 0.16$,

$p < 0.0001$, $n = 123$; $GC_{RNA23S} = 41.4 + 0.267\ T_{opt}$, $r^2 = 0.59$, $p < 0.0001$, $n = 18$. Again the effect remains robust in the comparative analysis (table 3). The contrasts for the largest of the analyses (16S) are shown in figure 4. The raw data for the tRNA analysis, however, suggest no effect: $GC_{tRNA} = 58.1 + 0.07\ T_{opt}$, $r^2 = 0.035$, $p = 0.25$, $n = 14$. Comparative analysis supports this (table 3). However, both the sample size and the sequences are very small and so firm conclusions should not be drawn.

4. DISCUSSION

The thermal adaptation hypothesis predicts that large positive contrasts in temperature should be matched by large positive differences in GC content, i.e. a positive slope of the regression of contrasts. Early data and analysis of completely sequenced genomes suggested that this may be true if one examines GC_3. Our results fail to provide any strong support for this pattern. Importantly, in both Eubacteria and Archea variation in GC_3 behaves like variation in $GC_{genomic}$, rather than like GC_{RNA}, in that GC_3 shows no covariation with temperature while GC_{RNA} shows a strong dependence. These results demonstrate that, at least within prokaryotes, GC content at both freely evolving and constrained sites within protein-coding sequences cannot be considered an adaptation to the thermal environment.

The GC_{RNA} results lend credence to the view that GC content and temperature may be related. With this aside, our results cannot be considered supportive of the model that supposes that elevated G and C contents, especially as witnessed at codon third sites, in genomes of plants, mammals and birds are an adaptation to higher temperatures (Bernardi 2000; Bernardi & Bernardi 1986; Salinas *et al*. 1988). This model has also failed to receive support from recent analysis of GC content of a few genes in a crocodile and a turtle that, while cold-blooded, appear to have GC contents similar to those in warm-blooded species (Hughes *et al*. 1999). These data should not, however, be over interpreted as the sample size is very small (Bernardi 2000) and the thermal physiology of the species concerned is uncertain.

The present result is also not definitive evidence against the homeothermy hypothesis as applied to non-prokaryotic taxa for, as Bernardi (2000) argues, it may be the case that some organisms solve the problem one way, while others do it differently. Indeed, if one supposes that archeal histone-like proteins are part of their defence against thermal degradation, then clearly they and Eubacteria have found different solutions to the same problem, at least to some extent. Hence, these results, while not supportive of the homeothermy hypothesis as applied to mammals and birds, do not falsify it. Nonetheless given that eukaryotes also have chromatin, it is unclear that they need necessarily have a problem with DNA thermal degradation.

We thank Mike Danson for advice on optimal growth temperatures and for two referees for helpful comments. This work was performed while A.R.M. was on a Nuffield Placement. L.D.H. is funded by The Royal Society.

REFERENCES

Bernardi, G. 2000 Isochores and the evolutionary genomics of vertebrates. *Gene* **241**, 3–17.

Bernardi, G. & Bernardi, G. 1986 Compositional constraints and genome evolution. *J. Mol. Evol.* **24**, 1–11.

Bruno, W. J., Socci, N. D. & Halpern, A. L. 2000 Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol. Biol. Evol.* **17**, 189–197.

Clay, O., Caccio, S., Zoubak, S., Mouchiroud, D. & Bernardi, G. 1996 Human coding and noncoding DNA: compositional correlations. *Mol. Phyl. Evol.* **5**, 2–12.

Galtier, N. & Lobry, J. R. 1997 Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J. Mol. Evol.* **44**, 632–636.

Grayling, R. A., Sandman, K. & Reeve, J. N. 1996 Histones and chromatin structure in hyperthermophilic Archaea. *FEMS Microbiol. Rev.* **18**, 203–213.

Harvey, P. H. & Purvis, A. 1991 Comparative methods for explaining adaptations. *Nature* **351**, 619–624.

Hughes, S., Zelus, D. & Mouchiroud, D. 1999 Warm-blooded isochore structure in Nile crocodile and turtle. *Mol. Biol. Evol.* **16**, 1521–1527.

Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T. & Oshima, T. 1984 High guanine plus cytosine content in the 3rd letter of codons of an extreme thermophile: DNA-sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. *J. Biol. Chem.* **259**, 2956–2960.

Maidak, B. L. (and 11 others) 2000 The RDP (Ribosomal Database Project) continues. *Nucl. Acids Res.* **28**, 173–174.

Muto, A. & Osawa, S. 1987 The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl Acad. Sci. USA* **84**, 166–169.

Purvis, A. & Rambaut, A. 1995 Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. *Computer Appl. Biosci.* **11**, 247–251.

Salinas, J., Matassi, G., Montero, L. M. & Bernardi, G. 1988 Compositional compartmentalization and compositional patterns in the nuclear genomes of plants. *Nucl. Acids Res.* **16**, 4269–4285.

Wada, A. & Suyama, A. 1986 Local stability of DNA and RNA secondary structure and its relation to biological functions. *Prog. Biophys. Mol. Biol.* **47**, 113–157.

Winter, G., Koch, G. L. E., Hartley, B. S. & Barker, D. G. 1983 The amino acid sequence of the tyrosyl transfer RNA synthetase from *Bacillus stearothermophilus*. *Eur. J. Biochem.* **132**, 383–387.

An electronic appendix to this paper can be found at http://www.pubs.royalsoc.ac.uk.